Supplemental Methods

Clinical Symptoms

The SDQ is an empirically valid epidemiological survey (1) suitable for use in community-recruited populations, and is sufficiently sensitive to detect clinical problems where they are present (2). Parent-youth rated agreement in SDQ symptom severity for the three symptom domains of interest (conduct, emotional and hyperactivity-inattention) while not as strong as Parent-Teacher ratings, tend to be in the moderate range and remain preferable to relying simply on one informant. Goodman et al (3) compared the concordance between Parent-Youth rated symptoms in adolescents (mean age 13 years) and found moderate agreement between both parties for conduct (correlation coefficient 0.36), hyperactivity (0.29) and emotional problems (0.52). A more recent investigation by Muris et al (4) supported this finding, reporting Parent-Youth inter-rater correlation coefficient of 0.31 for conduct, 0.42 for hyperactivity and 0.43 for emotional symptoms. We found Parent-Youth inter-rater agreement in the IMAGEN sample was also in the moderate range; emotional problems ($r=0.37$, $p<0.001$, $n=1288$), conduct symptoms ($r=0.32$, $p<0.001$, $n=1288$) and hyperactivity-inattention symptoms ($r=0.44$, $p<0.001$, $n=1288$). This study required a single symptom count variable for each domain of interest, therefore we compared the ratings of parents and youths for each item within each symptom subscale, and used whichever rating merited the more severe value in terms of symptom impairment. We did not average these scores as both scores were not always available for every participant. Furthermore, parent- and youth-rated scores are generally moderately correlated (conduct problems: $r = 0.41$; hyperactivity/inattention: $r = 0.41$; emotional symptoms: $r = 0.37$ (1)). These item level scores were summed and a combined symptom count score calculated for use in this study.

Table S1. SDQ frequency table by score

Group Classification	Conduct Disorder (n)	Group Classification	Hyperactivity (n)	Group Classification	Emotional (n)
Normal 0-2	744	Normal 0-5	929	Normal 0-3	941
Slightly Elevated 3-4	398	Slightly Elevated 6	150	Slightly Raised 4	141
High 5	86	High 7	103	High 5-6	116
Very High/Abnormal 6-10	60	Very High/Abnormal 8-10	106	Very high/Abnormal 7-10	90

Stressful Life Events

A self-report measure (5) was used to index events that had occurred during the lifespan and in the 12 months prior to the assessment using Psytools software (Delosis Ltd, London, UK) via its internet-based platform. Twenty events were classed as stressful based on the reports of IMAGEN participants (N=1239) who had experienced the event and rated it as distressing ('unhappy' or 'very unhappy'). These were; parents divorced, family had money problems, parents argued or fought, parent abused alcohol, brother or sister moved out, death in the family, family accident/illness, serious accident/illness, got or gave a sexually transmitted disease (STD), broke up with a boyfriend/girlfriend, in trouble with the law, stole something valuable, got in trouble at school, ran away from home, changed schools, family moved, face broke out with pimples, thought about suicide, got poor grades in school, gained a lot of weight. Sixteen events were consistently rated as positive ('happy' or 'very happy') and were classed as positive life events; these are not considered here.

A stressful life event frequency (SLEF) score was calculated based on the number of negative events that had occurred during the 12 months prior to the assessment only to avoid recall inaccuracy. The questionnaire was administered in such a way that events that had occurred in the previous 12 months could only be counted once.

fMRI Measurement and Processing

Structural and functional MRI data were acquired across all IMAGEN sites with 3T MRI scanners (Siemens, Philips, General Electric & Bruker). All sites used the same scanning protocol; high-resolution T1-weighted 3-dimensional structural images were acquired for anatomical localisation and registration with the functional time series. Blood oxygen-level dependent (BOLD) functional images were acquired with a gradient-echo, echo-planar imaging (EPI) sequence. In this task, 160 volumes were acquired per person, containing 40 slices (2.4mm slice thickness, 1mm gap) aligned to the anterior commission/posterior commission line and the echo-time optimised (TE 30ms; TR 2200ms) to reliably image subcortical regions. Data were pre-processed centrally (Neurospin, CEA) using SPM8 (Statistical Parametric Mapping, http://www.fil.ion.ucl.ac.uk/spm/).

Functional image data were corrected for movement (realigned to the first volume), slice timing, non-linear warping of each EPI to a custom EPI template and were smoothed with a 5mm full-width half maximum

Gaussian filter. Estimated movement parameters were added to each design matrix in the form of 18 additional columns (3 translations, 3 rotations, 3 quadratic and 3 cubic translations, and each 3 translations with a shift of ±1 TR). Twenty individuals were excluded due to poor realignment. Normalised and smoothed single-subject contrast images were taken forward to second level random effects analysis. Three contrasts were investigated; angry faces vs. control, neutral faces vs. control and angry faces vs. neutral faces (see Figure S1). Our target region of interest (amygdala) was significantly activated in two of the contrasts; anger vs. baseline (Table 3) and neutral vs. baseline (Table 4), but not in the anger vs. neutral contrast (see Table 5).

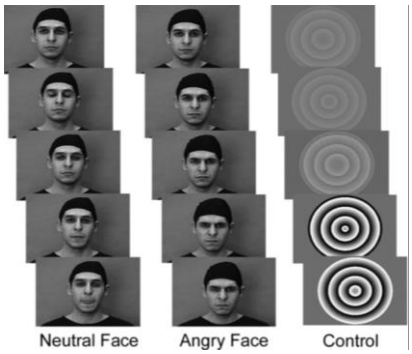

Figure S1. Emotional reactivity task showing neutral, angry, and control stimuli.

Using a separate sample of IMAGEN participants (N=326), not included in the experimental sample for reasons such as incomplete phenotypic data, we defined a functional amygdala region of interest (ROI) using a second-level ANOVA to generate a contrast giving equal weighting to angry and neutral conditions, whilst subtracting the control condition (0.5 0.5 -1.0). There was bilateral activation in the left amygdala: left amygdala maximal voxel MNI coordinates x = -18 y = -7 z = -14, F_{1975} = *13.71, β = 0.55, P<0.001;* right amygdala maximal voxel MNI coordinates x = 21 y = -7 z = -14, *k* = 2141, F_{1975} = *17.99, β = 0.60, P<0.001*. Using these MNI coordinates we created an amygdala ROI with a 8mm sphere. We confirmed there was significant activation in the ROI in the experimental sample (N=1288) in the angry faces vs. control contrast (Left: x = -18 y = -7 z = -14, *k*=80, T_{1287}=35.44, β= 0.609, p<.05 FWE; Right: x = 21 y = -7 z = -14, *k* = 81, T_{1287}=41.83, β= 0.570, p<.05 FWE; see Figure 2) and in the neutral faces vs. control contrast (Left: x = -18 y = -7 z = -14, *k* = 80, T_{1287} = 35.52, β = 0.558 , p<.05(FWE); Right: x = 21 y = -7 z = -14 , *k* = 81, T_{1287} = 46.74, β = 0.591, p<.05(FWE); see Figure 3).

Summarised beta values were extracted using MarsBaR [(6); http://marsbar.sourceforge.net/], and exported for analysis in SPSS.

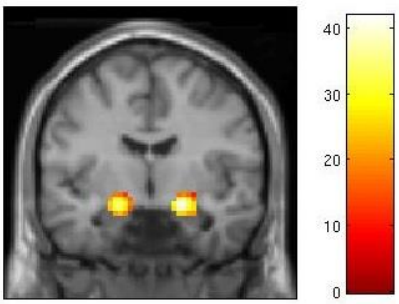

Figure S2a. Coronal view of a T1-weighted structural brain image illustrating bilateral amygdala region of interest in the angry faces vs. control contrast.

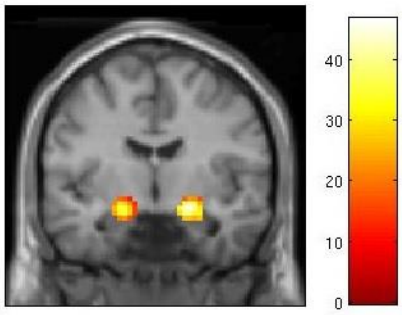

Figure S2b. Coronal view of a T1-weighted structural brain image illustrating bilateral amygdala region of interest in the neutral faces vs. control contrast.

PPI analyses

Generalized PPI (gPPI) regression analyses were carried out via the standard procedures of the SPM-based CONN toolbox ((7); http://www.nitrc.org/projects/conn) using the same amygdala ROI as our *a priori* seed region. gPPI also includes interaction factors from all conditions simultaneously in the estimation model to better account for between-condition overlaps. Functional data were band-pass filtered ($0.008 < f < 0.1$ Hz) to remove physiological noise. BOLD signal from white matter and CSF (5 dimensions each derived from the T1) were regressed out of the signal. Pre-processing-derived movement parameters and general task-related effects (first derivative) were also removed. Analyses were conducted on the subsequent BOLD signal. Analyses were

thresholded at p<0.001 (voxel-level uncorrected) and statistically significant clusters were reported at p<0.05 (family-wise error corrected). To adjust for multiple testing we applied the same Bonferroni correction (p<.05/12). We also ran non-parametric statistics (n=5000 permutations).

Table S2a. Brain regions significantly activated in the full sample of adolescents (N=1288) in the Angry Faces vs. Neutral Faces contrast. Whole-brain voxel-wise random effects analysis (Height threshold T=4.85, $p_{FWE\text{-}corrected}$<0.05, extent threshold, *k*=10 voxels). We report only T-scores for the sub-peaks of each significant cluster.

Anatomical Region	Left/Right	Coordinates of Peak Activation[a] (MNI)	*t*-score	β-value	Cluster Size (*k*)	*p-value FWE-corrected*
Anterior Cingulate Cortex		0 47 1	16.38	.36	6434	<.0001
Middle Cingulate Cortex		0 -22 34	12.74			
Insula	Left	-33 17 -14	9.76	.16	121	$7.44x10^{-15}$
Middle Orbital Gyrus	Left	-30 44 -11	5.68			
Inferior Frontal Gyrus	Left	-30 29 -14	5.42			
Inferior Parietal Lobule	Right	48 -58 49	9.04	.13	365	<.0001
Angular Gyrus	Right	51 -61 37	8.79			
Insula	Right	30 20 -11	7.99	.11	38	$5.99x10^{-8}$
Superior Frontal Gyrus	Left	-21 41 34	7.87	.09	228	<.0001
Inferior Temporal Gyrus	Left	-48 8 -35	7.74	.10	135	$7.77x10^{-16}$
Medial Temporal Pole	Left	-39 23 -35	6.11			
Middle Temporal Gyrus	Left	-63 -13 -20	6.11			
Middle Temporal Gyrus	Right	63 -25 -17	7.61	.09	41	$2.95x10^{-8}$
Lingual gyrus	Right	3 -79 -8	6.91	.14	40	$3.72x10^{-8}$
Sub-lobar, Extra-Nuclear	Left	-12 -1 -11	6.79	.11	31	$3.39x10^{-7}$

Caudate	Left	-12 11 -11	5.47			
Postcentral Gyrus	Left	-39 -19 40	6.48	.09	11	$1.30x10^{-4}$
Middle Frontal Gyrus	Right	33 38 46	6.48	.11	102	$1.79x10^{-13}$
Superior Frontal Gyrus	Right	27 44 40	5.85			
Olfactory cortex		0 17 -5	6.42	.12	28	$7.42x10^{-7}$

FWE, family-wise error; MNI, Montreal Neurological Institute; [a] Coordinates refer to the voxel with the maximum signal intensity

Table S2b. Demographic information for the amygdala region of interest test sample

	Verbal IQ	SES	Handedness	PDS	Parent CP	Youth CP
Missing Data (N)	67	79	6	16	7	6
Mean	105.266409	0.83	-	3.56	1.69	2.13
SD	17.666442	1.155	-	0.729	1.728	1.556
Min-Max	50-152	0-6	-	0-5	0-8	0-8
Pass/Fail QC (N)					7	5

Supplemental Results

Comorbidity

We examined the correlations between the symptoms, considering their comorbidity. Hyperactivity and conduct symptoms had the strongest association ($r = 0.445$, $p = 1.37 \times 10^{-63}$). Weaker correlations existed between emotion and the other two domains: (with hyperactivity, $r = 0.235$, $p = 1.20 \times 10^{-17}$; with conduct symptoms, $r = 0.226$, $p = 2.29 \times 10^{-16}$).

Site Differences

Considering the multisite nature of the study, engendering demographic differences, as well as scanner manufacturer and head coil differences, we explored whether the effect of site affected our main ROI results. We conducted a meta-analysis using the betas and standard errors from the regression models conducted separately for each site. We found that the sites have similar non-zero effect sizes and show overall significance in the models. This is true when examining 1) the interaction of conduct symptoms and stressful life events on right amygdala activation during the angry versus control contrast ($p=0.000573$); 2) the interaction of conduct symptoms and stressful life events on left amygdala activation in the neutral versus control contrast ($p=0.0111$); and 3) the interaction of hyperactivity symptoms and stressful life events on left amygdala activation in the angry versus control contrast ($p=0.0387$). Based on these data, we conclude that our model results are stable across sites.

Table S3a. Descriptive statistics by site

Site		Pubertal Status	Verbal IQ	SES	Psychosocial Stress Frequency	Age	Conduct Symptoms	Hyperactivity Symptoms	Emotion Symptoms
London	Mean	3.67	112.39	.75	3.30	14.3393	2.3971	4.83	2.37
	N (F=111 M=93)	204	204	204	204	204	204	204	204
	Std. Deviation	.712	12.209	1.021	1.998	.37863	1.53584	2.066	2.388
Nottingham	Mean	3.73	103.78	.67	3.22	14.4564	2.3710	4.74	2.38
	N (F=103 M=83)	186	186	186	186	186	186	186	186
	Std. Deviation	.670	12.303	1.054	2.210	.27828	1.73268	2.278	2.249
Dublin	Mean	3.57	109.97	.77	3.15	14.3599	2.2979	4.51	2.53
	N (F=51 M=43)	94	94	94	94	94	94	94	94
	Std. Deviation	.664	13.333	1.140	1.866	.29968	1.55079	2.368	2.327
Berlin	Mean	3.65	114.82	.85	3.78	14.4800	2.6727	4.44	2.60
	N (F=65 M=45)	110	110	110	110	110	110	110	110
	Std. Deviation	.698	14.070	1.340	2.065	.32869	1.69786	1.947	2.585
Mannheim	Mean	3.64	112.07	.79	3.67	14.3182	2.5367	4.10	2.33
	N (F=118 M=100)	218	218	218	218	218	218	218	218
	Std. Deviation	.631	15.161	1.074	2.206	.39638	1.51217	2.067	2.317

Hamburg	Mean	3.69	105.69	.53	3.37	14.4527	2.3772	3.75	1.69
	N (F=100 M=67)	167	167	167	167	167	167	167	167
	Std. Deviation	.685	14.099	1.017	2.064	.49007	1.39542	2.108	2.021
Paris	Mean	3.51	122.03	.37	3.06	14.3683	2.8365	4.33	2.08
	N (F=104 M=104)	208	208	208	208	208	208	208	208
	Std. Deviation	.798	15.017	.762	2.016	.45795	1.69461	2.205	2.145
Dresden	Mean	3.67	114.58	.73	3.25	14.5461	2.2178	3.83	2.16
	N (F=53 M=48)	101	101	101	101	101	101	101	101
	Std. Deviation	.694	12.148	1.139	2.037	.38121	1.36091	2.084	2.411
Total	Mean	3.64	111.98	.66	3.35	14.4018	2.4876	4.34	2.24
	N (F=705 M=583)	1288	1288	1288	1288	1288	1288	1288	1288
	Std. Deviation	.700	14.823	1.057	2.084	.39612	1.58121	2.167	2.297

Table S3b. Table of descriptive statistic site differences – covariates and symptoms

				Post-Hoc					
	ANOVA							95% CI	
Measure	**F**	**DF**	**P**	**Direction**	**mean diff**	**SE**	**P**	**lower**	**upper**
PDS	1.76	(7, 1280)	.091	-					
VIQ[1]	31.982	(7, 1280)	$4.34x10^{-41}$	Lon>Nott	8.61	1.24	$4.99x10^{-10}$	4.82	12.40
				Lon>Ham	6.70	1.39	.000056	2.47	10.93
				Dub>Nott	6.19	1.65	.005520	1.14	11.23
				Berl>Nott	11.04	1.62	$2.67x10^{-9}$	6.09	15.99
				Berl>Ham	9.12	1.73	.000008	3.83	14.41
				Mann>Nott	8.29	1.37	$8.54x10^{-8}$	4.12	12.45
				Mann>Ham	6.37	1.50	.000695	1.81	10.94
				Paris>Lon	9.64	1.35	$1.16x10^{-10}$	5.53	13.74
				Paris>Nott	18.25	1.38	$5.44x10^{-13}$	14.05	22.45
				Paris>Dub	12.06	1.73	$1.10X10^{-9}$	6.78	17.34
				Paris>Berl	7.21	1.70	.0000812	2.02	12.40
				Paris>Mann	9.96	1.46	$9.35x10^{-10}$	5.51	14.41
				Paris>Ham	16.33	1.51	$5.49x10^{-13}$	11.74	20.93
				Paris>Dres	7.45	1.60	.000137	2.57	12.32
				Dres>Nott	10.81	1.51	3.77^{-10}	6.19	15.42
				Dres>Ham	8.89	1.63	.000003	3.91	13.87
SES[1]	4.158	(7, 1280)	.000155	Lon>Paris	.39	.089	.000406	.12	.66
				Nott>Paris	.30	.094	.030	.02	.59
				Dub>Paris	.40	.129	.046	.00	.80
				Berl>Paris	.49	.14	.012	.06	.91
				Man>Paris	.42	.090	.000092	.15	.70
Age Years[1]	6.065	(7, 1280)	$5.57x10^{-7}$	Nott>Lon	.12	.033	.012	.015	.219
				Nott>Mann	.034	.033	.001298	.036	.24
				Berl>Lon	.141	.041	.016	.015	.266
				Berl>Mann	.162	.041	.002802	.036	.288

				Dre>Lon	.21	.046	.000347	.065	.348
				Dre>Dub	.186	.049	.004649	.036	.336
				Dre>Mann	.228	.046	.000052	.086	.370
				Dre>Par	.178	.049	.009314	.0265	.3291
Conduct[1]	2.687	(7, 1280)	. 009162	Par>Dre	.62	.179	.015	.0705	1.166
Hyper[2]	5.641	(7, 1280)	.000002	Lon>Mann	.74	.209	.012	.08	1.39
				Lon>Ham	1.085	.223	.000038	.39	1.78
				Lon>Dre	1.00	.260	.003527	.19	1.82
				Nott>Ham	.993	.228	.000405	.28	1.71
				Nott>Dres	.910	.265	.017	.08	1.74
Emotion[1]	2.400	(7, 1280)	.019259	Ber>Ham	.911	.292	.042	.02	1.81

[1]Post-Hoc Test = Games Howell [2]Post-Hoc Test = Bonferroni

fMRI Analysis Results

ROI results controlling for substance use:

We also ran the regression analyses controlling for substance use (defined in the Methods section). Including substance use in the regression models did not change our results. The interactions remained significant with respect to correction for multiple comparisons (angry vs. control contrast, conduct by stress in right amygdala: $p_{\text{family-wise error-corrected}}=0.003$, $p_{\text{Bonferroni-corrected}}=0.036$; neutral vs. control contrast, conduct by stress in left amygdala: $p_{\text{family-wise error-corrected}}=0.001$, $p_{\text{Bonferroni-corrected}}=0.012$; angry vs. control contrast, hyperactivity by stress in left amygdala: $p_{\text{family-wise error-corrected}}=0.002$, $p_{\text{Bonferroni-corrected}}=0.024$). There was only a main effect of drug use ($p_{\text{family-wise error-corrected}}=0.025$) for the conduct by stress interaction on left amygdala activation in the angry vs. control contrast that does not survive multiple comparison correction ($p_{\text{Bonferroni-corrected}}=0.3$). The emotional symptom regression models were still non-significant when controlling for substance use.

Table S3c. Brain regions significantly activated in adolescents in the Angry Faces vs. Control contrast as a function of conduct, hyperactivity/inattention problems and stressful life events. Whole-brain voxel-wise regression analysis (height threshold T=3.10, $p_{uncorrected}$<0.001, extent threshold, *k*=10 voxels), p-values are reported at family-wise error-corrected cluster level.

Region	Left/Right	Coordinates of Peak Activation[a] (MNI)	*t*	β	Cluster Size (*k*)	p
Main effect: Conduct Problems						
Superior parietal lobule and precuneus BA7	Right	24, -73, 49[a]	4.42	.05	48	.015
Postcentral gyrus	Right	33, -52, 73[a]	4.18	.04	51	.011
Interaction effect: Conduct Problems x Psychosocial Stress Frequency						
Region	**Left/Right**	**Coordinates of Peak Activation[a] (MNI)**	***t***	**β**	**Cluster Size (*k*)**	**p**
Middle Temporal Gyrus and Superior Temporal Gyrus	Right	69, -40, 7[a]	5.41	.02	83	$5.72x10^{-4}$
Superior Temporal Gyrus and Middle Temporal Gyrus	Right	42, -10, -14[a]	4.85	.02	137	$2.15x10^{-5}$
Thalamus and Pulvinar	Left	-12, -28, 13[a]	4.85	.02	124	$2.15x10^{-5}$
Insula, Rolandic Operculum, BA13 and Superior Temporal Gyrus	Right	36, -22, 19[a]	4.78	.02	65	4.78
Anterior Cingulate Cortex and Medial Frontal Gyrus	Left	-6, 47, 7[a]	4.72	.02	191	$2.03x10^{-7}$
Superior Frontal Gyrus and Middle Frontal Gyrus	Right	30, 62, 19[a]	4.69	.03	62	$3.8x10^{-3}$
Extra-Nuclear, Sub-lobar	Right	12, -31, 10[a]	4.53	.02	47	.017
Inferior Frontal Gyrus	Right	60, 23, 10[a]	4.42	.01	42	.028
Middle Cingulate Cortex	Right	3, 5, 37[a]	4.38	.02	128	$1.60x10^{-5}$
Middle Frontal Gyrus	Right	36, 38, 46[a]	4.30	.02	74	$1.3x10^{-3}$

Thalamus, Medial Globus Pallidus	Right	15, -16, 1[a]	4.28	.01	71	$1.6x10^{-3}$
Inferior Frontal Gyrus	Left	-51, 14, 4[a]	3.98	.02	39	.038
Interaction effect: Hyperactivity/Inattention Problems x Psychosocial Stress Frequency						
Region	**Left/Right**	**Coordinates of Peak Activation[a] (MNI)**	***t***	β	**Cluster Size (*k*)**	**p**
Middle Cingulate Cortex BA24, Anterior Cingulate Cortex	Left	-3, 2, 37[a]	7.09	.01	49	.014

MNI, Montreal Neurological Institute; [a]Coordinates refer to the voxel with the maximum signal intensity.

References

1. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. **J Am Acad Child Adolesc Psychiatry**. 2001;40(11):1337-45.

2. Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. **Br J Psychiatry**. 2000;177:534-9.

3. Goodman R, Meltzer H, Bailey V. The Strengths and Difficulties Questionnaire: a pilot study on the validity of the self-report version. **Eur Child Adolesc Psychiatry**. 1998;7(3):125-30.

4. Muris P, Meesters C, van den Berg F. The Strengths and Difficulties Questionnaire (SDQ)--further evidence for its reliability and validity in a community sample of Dutch children and adolescents. **Eur Child Adolesc Psychiatry**. 2003;12(1):1-8.

5. Newcomb MD, Huba GJ, Bentler PM. A multidimensional assessment of stressful life events among adolescents. **J Health Soc Behav**. 1981;22:400-15.

6. Brett M, Anton JL, Valabregue R, Poline JB. Region of interest analysis using ans SPM toolbox. 8th International Conference on Functional Mapping of the Human Brain; June 2-6.; Sendai, Japan 2002.

7. Whitfield-Gabrieli S, Nieto-Castanon A. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. **Brain Connect**. 2012;2(3):125-41.